DOCUMENT RESUME

ED 460 612 FL 025 295

AUTHOR Nakamura, Yuji
TITLE An Examination of English Speaking Tests and Research on English Speaking Ability.
PUB DATE 1994-03-20
NOTE 26p.
PUB TYPE Reports - Research (143)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Doctoral Dissertations; Elementary Secondary Education; *English (Second Language); Foreign Countries; Higher Education; *Language Proficiency; *Language Research; *Language Tests; *Oral Language; Second Language Instruction; Second Language Learning; Test Reliability; Test Validity; Testing
IDENTIFIERS Japan; *Japanese Speaking

ABSTRACT

This paper examines both overseas and domestic tests of English speaking ability from the viewpoint of the crucial testing elements such as definition of speaking ability, validity, reliability, and practicality. The paper points out problems to be solved and proposes suggestions for constructing an oral proficiency test in order to determine the detailed components of Japanese students' English speaking ability. The overseas research discussed includes the Test of English as Foreign Language Research Reports on Speaking Tests, Doctoral Dissertations, and other research on speaking ability. Six domestic research reports are examined, including three from the college level, two from the high school level, and one from the junior high school level. (VWL)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 460 612

An Examination of English Speaking Tests and Research on English Speaking Ability

Yuji Nakamura

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Yuji Nakamura

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

調布学園女子短期大学
紀要　第26号　抜刷
平成6年3月20日　発行

BEST COPY AVAILABLE

An Examination of English Speaking Tests and Research on English Speaking Ability

Yuji Nakamura

This paper examines both overseas and domestic tests of English speaking ability plus the research on the measurement of English speaking ability from the viewpoint of the crucial testing elements such as definition of speaking ability, validity, reliability and practicality. Eventually it points out problems to be solved and proposes suggestions for constructing an oral proficiency test in order to determine the detailed components of Japanese students' English speaking ability.

The overseas tests we shall examine are : ILR, TSE, ACTFL, PET, Pre-PET, ARELS, RSA, Ilyin Oral Interview Test, Upshur's Oral Communication Test and TOEIC. For the description of these tests, Davies and West (1989), and Alderson, Krahnke and Stansfield (1987) will occasionally be referred to. The domestic test we will examine is : STEP (EIKEN) Test (Ist Grade, Pre-1st Grade, 2nd Grade and 3rd Grade).

The overseas research we will discuss includes TOEFL Research Reports on Speaking Tests, Doctoral Dissertations, and other research on speaking ability. There are six domestic research reports we will examine (three from the college level, two from the high school level, one from the junior high school level).

1. Overseas Tests of Speaking Ability

1.1 Direct Speaking Tests

1) Cambridge Local Examinations

The University of Cambridge Local Examinations Syndicate conducts six Examinations in English as a Foreign Language and each has a speaking test in the form of an interview. They are listed in the order of descending difficulty as follows :

(1) Diploma of English Studies (DES)
(2) Certificate of Proficiency in English (CPE)
(3) Certificate in English for International Communication (CEIC)
(4) First Certificate in English (FCE)
(5) Preliminary English Test (PET)
(6) Pre-Preliminary English Test (Pre-PET)

Among these six, the first four tests (DES, CPE, CEIC and FCE) are not relevant as a review work for the present test because their target population is more proficient than the candidates we are concerned with. For example, the FCE test which is the lowest level among these four is suitable for candidates whose TOEFL score is over 500. Also, it is relevant for those who wish to work in English as the medium of communication at a functional level, which is an unrealistic level to expect with the students the present research is concerned with.

Although both PET and Pre-PET have an interview test as a speaking test, they are administered differently. The PET interview test is conducted by a native speaker of English to all the candidates, whereas the Pre-PET interview test is conducted by Japanese raters to those candidates who passed the written test.

In the PET interview, the candidate must perform functions and tasks such as :

(1) self-introduction
(2) giving information about things
(3) giving directions
(4) talking about time
(5) role-playing in a task-based situation

A native English speaking rater evaluates the candidates' speaking ability through the 12-minute interaction. In order to keep reasonable inter-rater reliability, there is a training session for all the raters.

The Pre-PET test intentionally targets Japanese students with some consideration of the English teaching and learning situation in Japan. In the interview test, which is in fact a role play test, the candidate plays one role and asks questions of one Japanese rater and answers questions from

the other Japanese rater. The raters are in separate rooms and the candidate is given some preparation time after he/she receives a card description of the role. The whole process takes about six or seven minutes. This is a highly controlled role play and time-limited test.

In summary, among the six Cambridge tests, the speaking part of two tests (PET and Pre-PET) can be applicable to the assessment of Japanese students' speaking ability at the level we are concerned with.

However, both have drawbacks. PET is conducted and assessed by native speakers of English ; thus, Japanese teachers might have difficulty in applying the test in a classroom situation.

Pre-PET is easy to administer in the classroom situation since it is evaluated by Japanese raters. Nevertheless, there is some doubt if only the role play which is highly controlled can be used to adequately measure the students' speaking ability.

One of the biggest problems for the Pre-PET test is that the beginning level students who do not do well in the written tests cannot take even the lowest Pre-PET test, because it has a cut off point in the written test to select the candidates for the role play test.

On the other hand, the merit of Cambridge tests is its new insight toward the concept of speaking ability. A new idea of notions and functions from Communicative Competence is well presented in the tasks and elicitation techniques.

2) ILR, ACTFL and TOEIC

The Interagency Language Roundtable (ILR) Oral Proficiency Interview (OPI), which was formerly known as the Foreign Service Institute Oral Proficiency Interview (FSIOPI), is designed to measure oral English language skills of adolescents and educated adults.

The American Council on the Teaching of Foreign Languages (ACTFL) and Educational Testing Service (ETS) Proficiency Guidelines are derivative scales.

The difference between the ILR and the ACTFL is the scale of the assessment shown in Table 1

Table 1

Comparison between ILR Scale and ACTFL Scale

ACTFL SCALE	ILR SCALE
	5 Native or bilingual proficiency
	4+
Superior	Distinguished proficiency
	3+ Professional working proficiency
Advanced High	2+
	2
Advanced	Limited working proficiency
Intermediate High	1+
Intermediate Mid	1
Intermediate Low	Survival proficiency
Novice High	0+
Novice Mid	0
Novice Low	No practical proficiency

In each scale, the responses by the candidate are scored holistically by a trained interviewer within a 10 to 40 minutes period.

The interview part of Test of English for International Communication (TOEIC), which is designed to assess the English language speaking ability of adult non-native speakers of English in commerce and industry, was also developed by ETS. Thus, there can be a close similarity among these interview tests. However, the TOEIC interview test can be taken only by those who are among top two ranks in five (A B C D E) in the preliminary written multiple-choice test.

The weakest point of these three interview tests (or scales) is that

none of them (ACTFL, ILR, TOEIC) are designed to adequately discriminate among lower level students.

Practicality and reliability are problems as well. We need native speakers as raters, which is not realistic in our daily situations at school ; besides, the extensive training of raters is necessary to keep a high inter-rater reliability even though it is time-consuming and expensive.

Be that as it may, the interview test has a high face validity because it requires examinees to use spoken language.

In summary, the Oral Proficiency Interview tests (in the ILR scale, in the ACTFL scale and in the TOEIC test) are not relevant for the lower/intermediate Japanese students.

Although several problem areas of the interview tests of these three scales are pointed out, the ACTFL Guidelines offer the present researcher a valuable new dimension of speaking tests which includes the current theoretical aspects of Communicative Competence as shown in Table 2.

Table 2

Example for Superior Level

1. Fluency
2. Grammar
3. Pragmatic Competence
4. Pronunciation
5. Sociolinguistic Competence
6. Vocabulary

Among these six aspects, fluency, grammar, pronunciation and vocabulary are the ones traditionally included in the speaking test. However, Pragmatic Competence (confident use of various conversation management devices) and Sociolinguistic Competence (appropriate use of the major registers) are apparently derived from the idea of current theoretical framework of communicative competence.

3) Royal Society of Arts Examinations

The Royal Society of Arts Examination Board offers Examinations in the Communicative Use of English as a Foreign Language (RSACUFL) for adults non- native speakers primarily in Great Britain but also for ESL/EFL speakers outside of Great Britain at the basic, intermediate and advanced levels.

There are four components in the test and one of them is the Test of Oral Interaction. The aim of this interaction test is to be wholly authentic and to make language testing more communicative (Ogasawara 1987, Hargreaves 1987), which is a recommendable goal for the present test as well. The interaction test has three parts:

Part 1: interaction between the candidate and the examiner (interlocuter)

Part 2: interaction between candidates

Part 3: a report from the candidate to the examiner (interlocuter)

These three parts give the impression that this test stresses the authentic situation of the interaction.

Moreover, the description of the skills and the description of the levels of the test introduce new terms such as appropriacy, range, flexibility, size which are new aspects in the field of assessing oral proficiency.

However, for practical reasons such as the necessity of having native raters and the need of training of raters (native speakers) to keep a high reliability, this RSA test cannot be recommended to Japanese classroom teachers.

4) Ilyin Oral Interview Test and Upshur's Oral Communication Test

The Ilyin Oral Interview Test was developed by Donna Ilyin and is designed to assess oral proficiency in English in a controlled picture sequence situation and to provide diagnostic information on individual performance.

Upshur's Oral Communication Test is also conducted with pictures, and it is a very structured test similar to the Ilyin Test.

One of the significant features of these tests (Ilyin Test and Upshur's Test) is that they discriminate particularly well among lower or beginning level candidates. Furthermore, they are both easy to administer.

In addition, they have the merits of being a picture oriented test :

1) the picture gives something to talk about
2) the picture leads candidates into a narration

However, they have weaknesses as well. Tests users have concluded that raters should be trained more carefully in order to keep the inter-rater reliability high.

Upshur's test itself has weaknesses :

1) the time limitation to the candidate
2) no chronological or semantic relation among the four pictures

The Ilyin oral interview test is weak in that the test is an extreme case of overconcern for reliability but a denial of the natural flow of human interaction (Ogasawara 1987).

In summary, the picture device in these tests is recommendable especially to elicit responses from lower/ intermediate level students and to discriminate among them effectively. The present researcher has adopted the picture (visual-material) description technique in his research. Nevertheless, too much concern for reliability and too much control of the procedure will cause the flow of the interaction to be unnatural. Eventually, these problematic factors will lead the test task to become a non-authentic one. Therefore, we should remember that the more natural the language sample elicited, the greater the possibility of assessing the candidates' real speaking ability (Gonzalez 1990).

1.2 Semi-Direct Speaking Tests

1) TSE (Test of Spoken English)

The Test of Spoken English (TSE) was developed by Educational Testing Service (ETS) and is designed as a semidirect measurement of oral English skills of adults, especially graduate students and professionals whose native language is not English.

Like the TOEFL (Test of English as a Foreign Language) test, the TSE is administered at testing centers around the world by ETS or its representatives.

The test takes approximately 30 minutes and the candidates' answers

are recorded on tape. The questions are either printed or recorded. The questions used in the test are divided into seven sections in which test takers are asked :

(1) to answer questions about themselves
(2) to read printed passages aloud
(3) to complete partial sentences
(4) to construct a story from a series of pictures
(5) to answer questions about a single picture
(6) to answer questions on general topics
(7) to give a short presentation as if they were speaking to a group of students

The results (the test tapes) are scored by trained raters at ETS in four categories (fluency, pronunciation, grammar and overall comprehensibility) using two different scales (0.0 to 3.0 for fluency, pronunciation and grammar ; 0 to 300 for overall comprehensibility).

The TSE has strengths as follows :

(1) It has a greater face validity than paper and pencil tests (which are usually used in an academic setting to evaluate speaking ability or pronunciation).
(2) It has a greater administrative convenience than direct measurements such as ILROPI, ACTFLOPI.

However, it has to be pointed out that the TSE has three weaknesses as a speaking test for evaluating the speaking ability of Japanese students :

(1) The theoretical construct is not clear : what theoretical framework is the TSE based on or where can we see the concrete examples of sociolinguistic competence, etc.?
(2) Its cost is high.
(3) The target is so high that it cannot discriminate among lower level Japanese students : the TSE is appropriate only for advanced level candidates.

2) ARELS Oral Examinations

The Association of Recognized English Language Schools Examinations

Trust developed Oral Examinations (called ARELS Oral Examinations) at three levels--Preliminary/Junior, Higher and Diploma.

All of these examinations are conducted in the language laboratory. Candidates anywhere in the world listen to identical master tapes, and their responses are recorded on personal tapes, which are sent in for marking.

The two higher levels (Higher and Diploma) are only suitable for intermediate and higher level students, in other words, higher than the Council of Europe "Threshold" level (van Ek and Alexander 1975).

The preliminary level is approximately the Council of Europe "Waystage" level and suitable for lower/intermediate students.

The speaking part in these levels includes the following tasks :

(1) making appropriate responses
(2) reading aloud, text or part of dialogue
(3) narrating story from picture cues
(4) describing a picture
(5) giving a short talk on a chosen topic
(6) answering comprehension questions on a recorded text or dialogue
(7) summarizing a recorded passage
(8) sentence transformation and question formation
(9) interpreting stress and intonation patterns

There is always the criticism of a semi-direct speaking test like the ARELS Oral Examination which is conducted wholly on tape, of whether the interaction on tape can possibly be a natural communication from the viewpoint of authenticity.

However, the ARELS type semi- direct examination has many strengths :

(1) we can conduct the test under the same conditions at one time to a large number of people
(2) we can obtain a variety of a candidate's answers from various tasks and sub-tasks
(3) a variety of tasks can help construct a detailed definition of speaking ability
(4) we can obtain native speakers' criteria/evaluation of speaking

ability purely through the tape without any visual distractors

2. Existing Domestic Test of Speaking Ability—STEP Test

The Society for Testing English Proficiency (STEP) has developed tests for assessing Japanese candidates' English proficiency. There are six grades—1st Grade, Pre-1st Grade, 2nd Grade, 3rd Grade, 4th Grade and 5th Grade. The 1st to 3rd Grades have supplemental secondary speaking tests.

In the 1st Grade test, the speaking test is conducted through public speaking and two raters (including one native English speaker) assess the Content (content, accuracy, quantity) and Delivery (pronunciation, intonation, grammar and narration) of the speech.

In the Pre-1st secondary speaking test, the candidate should perform picture narration and question-answer activities. His or her pronunciation, fluency, vocabulary, grammar and content are evaluated in the picture narration. In the question-answer activities, the accuracy of the grammar and the content, the naturalness of the voice, etc., are evaluated holistically. All the performances are assessed by one rater (either a Japanese or a native speaker of English).

In the speaking test of 2nd and 3rd Grades, the candidate should demonstrate his/her "speaking" ability through reading aloud activity and question-answer activities. In the reading aloud activity, pronunciation, stress, juncture, speed and rhythm are evaluated, and in the question-answer activities, the accuracy of the content is evaluated by a Japanese rater.

The STEP test in their secondary speaking parts can be summarized from their weaknesses and strengths as follows:

Weaknesses

(1) Their theoretical framework is not clear.

(2) Although it is very popular, it cannot be said to be a standardized test from the statistical point of view.

(3) The question and answer activities in 2nd and 3rd grades are not

speaking tests.

(4) There are many students who cannot take the "speaking" test because they cannot pass the preliminary written (multiple choice) test. For example, the present researcher's students (approximately 300 in total) took the first written test of 2nd Grade and only 10% of them passed it. The test makers should clarify that the skills in the multiple choice test are reflecting the skills in the speaking test partially because the basic grammatical ability is common to all language skills. Otherwise, the results of the multiple choice test should not be a prerequisite for the speaking test.

(5) The criteria is established based on the English language ability of Japanese students ; therefore, it is doubtful whether a successful candidate of a speaking test can speak English well in a real life situation.

(6) Training for the examiners is not well established, which makes for questionable reliability. In fact, no scorer reliability figures have been released.

(7) Only one speaking ability task for each grade is not sufficient to evaluate a candidate's true speaking ability.

Strengths

(1) Since there are six grades, it is easy for the candidates to set their individual goal separately, and goal setting can be a good motivation to study English.

(2) Since the test is designed for Japanese students by Japanese test makers, there is little cultural gap in the questions.

3. Overseas Research on Speaking Ability

1) Stansfield, C.W. (1990)

Charles W.Stansfield (1990) introduces the simulated oral proficiency interview (SOPI) which is a type of semi-direct speaking test that models, as closely as is practical, the format of the oral proficiency interview (OPI) of ILR and ACTFL.

The SOPI is a tape-recorded test first developed by Clark and Li (1986), and named by Stansfield as such in 1986. In this test, a trained interviewer is not needed. Also, it can be administered to a group of examinees. Lastly, the tape is scored by a trained rater using the ACTFL/ILR scale.

The results showed the correlation between the SOPI and the OPI to be .93. In other words, a semi-direct speaking test can have a high correlation with a direct speaking test.

The SOPI demonstrates many practical and psychometric advantages over the OPI. Stansfield, taking into consideration the fact that the SOPI highly correlates with the OPI, claims that it seems safe to say that both the OPI and the SOPI measure the same abilities through the analysis of the results of many cases between them. Finally, he states that Clark's (1978: 48) characterization of semi-direct tests should be considered as "second-order substitutes for direct techniques."

2) Clark, J. L.D., and Swinton, S. S. (1980)

Clark and Swinton (1980) did research to determine the concurrent validity of the Test of Spoken English (TSE) in relation to the Foreign Service Institute (FSI) Oral Proficiency Interview (now ILROPI) by administering the two tests to 134 foreign teaching assistants at nine state universities.

The results indicate a high correlation (.79) between the TSE and FSI total scores. Thus, Clark and Swinton contend that the TSE can be considered a reasonable alternative to the FSI interview when it is not possible to carry out testing in a face to face setting.

Although this statement is milder and more reserved than that of Stansfield who strongly claims the equality of a direct test and a semi-direct test, Clark and Swinton's research gives another example of stressing the efficiency of a semi-direct speaking test.

3) Lowe, P. and Clifford, R. (1980)

Lowe and Clifford studied the correlation between the Recorded Oral

Proficiency Examination (ROPE) -- a semi-direct measure of overall proficiency and an oral interview--a version of the Foreign Service Institute oral interview.

The results showed a surprisingly high correlation (.90) between them. Their conclusion is that alternate test elicitation techniques like a semi-direct test which have satisfactory validity can be developed and used where a direct test such as an interview test is impossible.

Lowe and Clifford's research is also an example of stressing the efficiency of a semi-direct speaking test which can be a substitute for a direct test.

4) Oka's Review Work (1984)

According to Oka (1984), in addition to those existing tests of semi-direct speaking tests such as TSE, ARELS, some scholars have conducted research on the effectiveness of semi-direct speaking tests--Pimsleur (1961), Stack (1960). Stack's test can be considered practical for lower level students in terms of its length of the time and ease of scoring.

Research on a direct speaking test was conducted in the form of discrete point measurement (Clark 1972; Valette 1977; Pimsleur 1966), interview tests (Harris 1969, Heaton 1975) and speech making tests (Heaton 1975).

Furthermore, there have been communicative tests --role playing with picture stimuli (Ilyin 1972; Upshur 1971a).

5) Shohamy, E. (1983)

Shohamy (1983) dealt with the assessment of speaking ability through interviews and through reporting. What she found was that there was a difference in the students' speaking ability depending on the speech modes (interviews, question-answer, dialogues etc.). Her claim is that it is unfair to decide students' speaking ability only through the result of one task mode.

Shohamy's statement encouraged the present researcher to construct a large scale speaking test consisting of different task modes.

6) Sanders, S. L. (1981) and Riggenbach, H. R. (1989)

Sanders (1981) and Riggenbach (1989) dealt with speaking (or speeches) in their doctoral dissertations.

Sanders dealt with the techniques of conversational analysis to examine three features of the students' communicative competence (in the sense of oral proficiency) :

(1) ability to respond to different question types

(2) ability to produce expansions and accounts

(3) ability to perceive and focus topics

She obtained valuable insight concerning features to look for in scoring the oral proficiency interview through the application of conversational analysis. She further suggested a technique for examining student speech without the necessity of evaluating its correctness in the grammatical sense.

Her dissertation does not give much relevant information from the statistical point of view. However, the idea of these three features of communicative competence and the insistence of the focus on effective communication should be taken into account for the present test's scoring criteria.

Riggenbach tried to explore specific fluency—related features in the speech of six non-native subjects. Her findings suggest that fluency is a complex, high-order linguistic phenomenon, and that intuitive judgements about fluency level may take into account a wide range of linguistic phenomena.

It is not easy to surmise how fluency influences the raters' judgement ; nonetheless, Riggenbach's paper is a valuable study vis-a-vis reconsidering the definition of fluency in the present study.

4. Research in Japan on Speaking Ability

4.1 Studies on College Students

1) Ogasawara, Y. (1987)

Ogasawara tried to develop a workable set of scales to assess oral proficiency, especially at the lower levels of proficiency by following the model

of existing FSI Oral Interview (now ILROPI).

She took the stance that an interview would be most desirable for assessing oral proficiency, although she admits that the discrete-point test and the integrative test may be complementary and not mutually exclusive.

The structure of the interview is quite similar to the ACTFLOPI, starting with a warm-up, followed by a level check, a probe, and a wind-up. The inteviewer can use a picture telling technique for the lower level students. The test lasts for about ten minutes.

Ogasawara used 33 Japanese college students as subjects and 10 raters (five native English speakers and five Japanese teachers) scored the videotaped interviews using a 12 level scale.

What she found was an inter-rater reliability of (.972), which is surprisingly high in this type of oral production test.

The content validity was checked against the syllabus for the course, although the present researcher would like to know how the course syllabus was constructed from the theoretical framework of communicative competence.

Ogasawara's research is a valuable work to support the efficiency of an interview test ; nevertheless, the present researcher must ask the following questions in addition to her stated problems (level description and the importance of rater training) :

(1) Do we really need a 12-point scale?

(2) Is it possible to avoid the Halo effect in a face to face interaction?

(3) Are we sure what the speaking ability of this test is composed of?

(4) Is it practical to have an interview test which a Japanese classroom teacher conducts in the Japanese context where class size is of 30-40 students and administer each interview under the same conditions?

2) Morita, Y. (1987)

Morita (1987) contends that the teaching of speaking and eventually the evaluation of speaking ability should be conducted not only by native English speakers but also by Japanese teachers even at the college level.

After surveying Ilyin's Oral Interview Test, Upshur's Oral Communication Test, FSI Oral Interview and a semi-direct test using the language laboratory, he offered a pair-work speaking test using an information gap filling task in role playing, which is called "Aural-Oral Communication Ability Test". He reported the results obtained from his students

He stated the following merits and demerits of his test :

Merits

(1) It is feasible in a 30-40 student class.

(2) An information gap drill can be conducted in daily classroom activities to obtain high efficiency of the test, and eventually a high backwash effect can be expected.

Demerits

(1) Since the obtained scores are for a pair, it is difficult to know individual ability.

(2) If there is a large proficiency discrepancy between the two testees, the reliability is low.

(3) The quality of communication may not be measured.

Judging from these results, we cannot avoid facing the dilemma of a gap between practicality and reliability/ validity.

3) Nakasako, S. (1987) and Cantor, G. W. (1987)

Nakasako (1987) introduces a way of public speaking for the evaluation of Japanese students' English speaking ability. He suggests that teachers should try to improve their own evaluation system and to evaluate the students' oral proficiency as objectively as possible.

Cantor (1987), from the native English speaker's point of view as an English conversation teacher, recommends some sub-tasks for a speaking test (e. g., giving a short talk or report ; responding to situations ; discussions, interviews ; role plays). Furthermore he advises us to consider the following :

(1) whether to rate students on a number of different categories or simply on the basis of their overall fluency or communicative abil-

ity ; in other words, analytic assessment versus holistic assessment

(2) whether or not to record tests for grading purposes

(3) whether to use more than one person to grade tests

4.2 Studies on High School Students

1) Takeda, S. (1990)

Takeda (1990) evaluated Japanese high school students' (517 in total) speaking ability using a modified version of John Upshur's Productive Communication Test. He used only four test items.

From the practical point of view, his test is acceptable (it took four minutes per person). Also, face validity can be recognized since students used spoken language in the real situation.

However, we are not sure if his test was valid in other respects or reliable with only four test items and with only one task mode even if the number of his subjects was 517.

2) Sakai, S. (1991)

Sakai (1991) tried to develop a way of evaluating Japanese high school students' oral communication ability in a classroom setting.

He used a pair work test of an information gap filling task with role play for the test. His subjects were 14 (seven pairs) students. The notion of the pair work was selling/buying with 10 sub-test items.

His test is reasonable from the practical viewpoint and has face validity since the students did perform in English. Weaknesses of his test are :

(1) the number of notions (only one) is so restricted to investigate the theoretical application.

(2) the number of test modes (only one) is limited to assess students' diversified speaking ability.

4.3 Study on Junior High School Students

Uchiki (1991) proposed a way of assessing junior high school students' speaking ability by offering a combination of an informal test with a

formal test with the following considerations :

(1) the test can be conducted in a 40-student classroom setting

(2) the test should be a direct speaking test

In his test consisting of two tasks (a listening test and a speaking test) , there is peer evaluation of a speaking test for the informal test and a teacher's speech evaluation for the formal test.

The present researcher agrees with the idea that informal tests and formal tests should play a complementary role to grasp the fair evaluation of students' speaking ability. Nevertheless, he is not convinced with the validity of peer evaluation although he admits that peer evaluation facilitates students' involvement in the speaking activities from the viewpoint of mutual encouragement.

5. Summary and Conclusions

We examined tests of English speaking ability (overseas and domestic) and research on measurement of English speaking ability in terms of the crucial testing elements such as definition of speaking ability, validity, reliability, practicality and other related factors.

Firstly, we dealt with existing overseas speaking tests by seperating them into two categories (Direct Tests and Semi-Direct Tests) :

(1) Direct Tests

e. g., Cambridge Tests (PET, Pre-PET), ILROPI, ACTFLOPI

(2) Semi-Direct Tests

e. g., TSE, ARELS

Secondly, we discussed an existing domestic test (STEP Test) by focusing on the speaking part.

Thirdly we examined previous studies, not only overseas but also domestic, on speaking ability :

(1) Studies Overseas

e. g., Stansfield (1990) , Clark and Swinton (1980), Shohamy (1983), Sanders (1981), Riggenbach (1989)

(2) Studies in Japan

e. g., Ogasawara (1987), Morita (1987), Takeda (1990), Sakai

(1991)

This comprehensive study on existing speaking tests and the previous research provides us with the following findings :

(1) Existing overseas speaking tests cannot be directly employed for assessing Japanese students' speaking ability in a classroom situation because of their level of difficulty and lack of practicality in the Japanese classroom situation.

(2) There are almost no valid and reliable speaking tests available in a classroom setting to assess the lower and intermediate level students' speaking ability minutely.

(3) Few studies are concerned with the definition of the construct of speaking ability.

(4) Few Japanese scholars have conducted research on the influences of the speech modes (or the test tasks) on students' speaking performance.

(5) There are no large scale speaking tests developed in Japan which are based on native speakers' scoring standards that can be easily conducted by Japanese teachers.

Therefore we need to construct in Japan a speaking test by taking into account the following seven points :

1) The need, especially at the college level, to improve English speaking ability and the testing of English speaking ability

2) The recognition of the problems of using productive speaking tests with students who have been accustomed to passive tests such as true false tests, or multiple choice tests and are hesitant to demonstrate their speaking ability in English

3) The ambiguity of the definition of speaking ability in the framework of Communicative Competence

4) The lack of sufficient validity of available speaking tests

5) The lack of the reliability of present speaking tests

6) The inadequacy of traditional elicitation techniques for speaking tests

7) The unavailability of large scale speaking tests based on native

speakers' standards as well as experienced Japanese English teachers' standards for the Japanese teachers' use

Finally, it is hoped, we will be able to examine in the future the detailed components of Japanese students' English speaking ability from the viewpoint of Communicative Competence by exploring native speakers' rating system. Although Communicative Competence only has a 20 year track record, its creditability will be greatly enhanced when we make an objective test to evaluate Japanese students' English speaking ability. The uniqueness of this test is that it will be designed to be conducted in a classroom setting by Japanese teachers of English, easily, quickly, effectively and economically.

BIBLIOGRAPHY

ACTFL (The American Council on the Teaching of Foreign Languages). (1986). *ACTFL Proficiency Guidelines.* New York : The American Council on the Teaching of Foreign Languages.

ACTFL (The American Council on the Teaching of Foreign Languages). (1987). *ACTFL Japanese Proficiency Guidelines.* New York : The American Council on the Teaching of Foreign Languages.

ACTFL (The American Council on the Teaching of Foreign Languages). (1989). *The ACTFL Oral Proficiency Interview : Tester Training Manual.* New York : The American Council on the Teaching of Foreign Languages.

Alderson, J. C., Krahnke, K. J. and Stansfield, C. W. (Eds.). (1987). *Reviews of English Language Proficiency Tests.* Washington, DC : TESOL

American Psychological Association. (1985). *Standards for Educational and Psychological Testing.* Washington, DC : American Psychological Association.

ARELS Oral Examinations (Association of Recognized Language Schools Oral Examinations in Spoken English). Tony, T. (1987). Association of Recognized Language Schools Oral Examinations in Spoken English. In J. C. Alderson, K. J. Krahnke and C. W. Stansfield (Eds.), *Reviews of English Language Proficiency Tests* (pp.5-7). Washington, DC : TESOL.

ARELS Oral Examinations. Davies, S and West, R. (1989). *The Longman Guide to English Language Examinations* (pp.5-10). Longman.

Bachman, L. F. (1990). *Fundamental Considerations in Language Testing.* Oxford : Oxford University Press.

Cantor, G. W. (1987). Testing in the conversation classroom. *Eigogaku-Ronsetsu-Shiryo* (Papers in English Linguistics) 21, 683-689.

Clark, J. L. D. (1972). *Foreign-Language Testing Theory and Practice.* Philadelphia : The Center for Curriculum Development.

Clark, J. L. D. (1979). Direct vs. semidirect tests of speaking ability. In E.J.Briere and F.B.Hinofotis (Eds.), *Concepts in language testing : Some recent studies* (pp.35-49). Washington, DC : TESOL.

Clark, J. L. D. and Swinton, S. S. (1979). An Exploration of Speaking Proficiency Measures in the TOEFL Context. *TOEFL Research Reports. Report 4.* Princeton, NJ : Educational Testing Service.

Clark, J. L. D. and Swinton, S. S. (1980). The test of spoken English as a measure of communicative ability in English-medium instructional settings. *TOEFL Research Reports. Report 7.* Princeton, NJ : Educational Testing Service.

Clark, J. L. D. and Li, Y. (1986). *Development, validation, and dissemination of a proficiency-based test of speaking ability in Chinese and an associated assessment model for other less commonly taught languages.* Washington, DC : Center for Applied Linguistics.

Davies, S and West, R. (1989). *The Longman Guide to English Language Examinations.* Longman.

Dornyei, Z. and Thurrel, S. (1991). Strategic competence and how to teach it. *ELT Journal,* 45/1.

Duran, R. P., Canale, M., Penfield, J., Stansfield, C. W. and Liskin-Gasparro, J. (1985). *TOEFL from Comunicative Viewpoint of Language Proficiency : A Working Paper. TOEFL Research Reports Report 17.* Princeton, NJ : Educational Testing Service.

ETS (Educational Testing Service). (1985). Interagency Language Roundtable Skill Level Descriptions. In R. P. Duran, M. Canale, J. Penfield, C. W. Stansfield and J. E. Liskin-Gasparro) (Eds.), *TOEFL from a Communicative Viewpoint on Language Proficiency : A Working Paper.* TOEFL Research Reports Report 17 (Appendix E). Princeton, NJ : Educational Testing Service.

ETS (Educational Testing Service). (1990). *Test of Spoken English manual for score users.* Princeton, NJ : Educational Testing Service.

Gonzalez, R. D. (1987). Ilyin Oral Interview. In J. C. Alderson, K. J. Krahnke and C. W. Stansfield (Eds.), *Reviews of English Language Proficiency Tests* (pp.41-43). Washington, DC : TESOL.

Hargreaves, P. (1987). Royal Society of Arts (RSA) : Examinations in the Communicative Use of English as a Foreign Language. In J. C. Alderson, K. J.

Krahnke and C. W. Stansfield (Eds.), *Reviews of English Language Proficiency Tests* (pp.32-34). Washington, DC : TESOL.

Harris, D. P. (1969). *Testing English as a Second Language.* New York : McGraw-Hill.

Heaton, J. B. (1975). *Writing English Language Tests.* 2nd Impression (1989). London : Longman.

Ilyin Oral Interview. Ilyin, D. (1972). *Ilyin Oral Interview.* Rowley, Mass : Newbury House.

ILR (Interagency Language Roundtable). Pardee, L. (1987). Interagency Language Roundtable Oral Proficiency Interview. In J.C.Alderson, K. J. Krahnke and C. W. Stansfield (Eds.), *Reviews of English Language Proficiency Tests* (pp.43-47). Washington, DC : TESOL.

Lowe, P. and Clifford, R. (1980). Developing an indirect measure of overall oral proficiency. In J. R. Frith (Ed.), *Measuring Spoken Language Proficiency* (pp.31-39). Washington, DC : Georgetown University Press.

Morita, Y. (1987). Toward communicative efficiency. *Eigogaku Ronsetsu Shiryo* (Papers in English Linguistics) 21, 1-8.

Nakasako, S. (1987). Evaluation of public speaking for Japanese EFL/ESL students. *Eigogaku Ronsetsu Shiryo* (Papers in English Linguistics) 21, 87-92.

Negishi, M. (1990). Hanasu noryoku no hyoka (Assessment of speaking ability). *The Forum for English Language Teachers* (FELT) 6, 39-42. Tokyo : The Institute for Research in Language Teaching.

Ogasawara, Y. (1987). Interview Assessment Scale for Japanese University Students. *Studies in Foreign Language Teaching* 9, 3-78. Tsukuba University Center.

Oka, H. (Ed.). (1984). *Eigo no supiikingu* (English Speaking). Tokyo : Taishukan.

PET (Preliminary English Test). Fried-Booth, D. L. and Hashemi, L. (1988). *PET Practice Tests 1.* Cambridge University Press. Akagawa, Y. (Ed.) (1989), Japanese Edition. Tokyo : Kenkyusha.

Pimsleur, P. (1961). A French speaking proficiency test. *The French Review* 34, 5, 470-79.

Pimsleur, P. (1966). Speaking skill. In A. Valdman (Ed.), *Trends in Language Teaching* (pp.194-207). New York : McGraw-Hill.

Pre-PET (Pre-Preliminary English Test). Cambridge Examinations Development Unit. (1988). *Pre PET Test Book.* Tokyo : Kenkyusha.

Riggenbach, H. R. (1989). Nonnative fluency in dialogue versus monologue speech : A microanalytic approach. (Doctoral dissertation, University of California at Los Angeles). *Dissertation Abstracts International,* 50, 2039A.

RSA (Royal Society of Arts Examinations). Davies, S and West, R. (1989). *The Longman Guide to English Language Examinations* (pp.106-108). Longman.

Sakai, S. (1991). Eigono-oralukomyunikeshon-noryoku-o-sokuteisuru-tesutono-kaihatsu (A test development to evaluate oral communication ability in English). *STEP Bulletin* 3, 59-72. Tokyo : The Society for Testing English Proficiency.

Sanders, S. L. (1981). Negotiating Understanding in the ESL Oral Proficiency Interview. (Doctoral dissertation, University of South Carolina). *Dissertation Abstracts International*, 42, 5109A.

Shohamy, E. (1983). The stability of oral proficiency assessment on the oral interview testing procedures. *Language Learning* 33, 4, 527-40.

Stack, E. M. (1960). *The Language Laboratory and Modern Language Teaching*. New York : Oxford University Press.

Stansfield, C. W. (1990). An evaluation of simulated oral proficiency interviews as measures of spoken language proficiency. In J. E. Alatis (Ed.), *Georgetown University Round Table Language and Linguistics 1990* (pp.228-34). Washington, DC : Georgetown University Press.

STEP (The Society for Testing English Proficiency). (1992). *STEP Bulletin* 4, 158-59. Tokyo : The Society for Testing English Proficiency.

Takeda, S. (1990). Eigoniyoru-komyunikeishon-no-sokuteini-kansuru-kenkyu (A study on the evaluation of communication ability in English). *STEP Bulletin* 2, 71-85. Tokyo : The Society for Testing English Proficiency.

TOEIC (Test of English for International Communication). Perkins, K. (1987). Test of English for International Communication. In J. C. Alderson, K. J. Krahnke and C. W. Stansfield (Eds.), *Reviews of English Language Proficiency Tests* (pp.81-83). Washington, DC : TESOL.

TSE (Test of Spoken English). Bailey, K. (1987). Test of Spoken English. In J. C. Alderson, K. J. Krahnke and C. W. Stansfield (Eds.), *Reviews of English Language Proficiency Tests* (pp.84-86). Washington, DC : TESOL.

TSE (Test of Spoken English). (1990). *TSE Manual for Score Users*. Princeton, NJ : Educational Testing Service.

Uchiki, K. (1991). Supiikinguniokeru-hatuswanoryoku-to-taionoryoku-no-ikuseito-hyoka (How to assist and evaluate students' speaking ability-productive ability and interactive ability). *STEP Bulletin* 3, 29-45. Tokyo : The Society for Testing English Proficiency.

University of Cambridge Local Examinations Syndicate. (1989). Examination summaries. In S. Davies and R. West (Eds.), *The Longman Guide to English Language Examinations* (pp.114-128). Longman.

Upshur's Oral Communication Test. Upshur, J. A. (1971a). Objective evaluation of oral proficiency in the ESOL classroom. *TESOL Quarterly* 5, 1, 47-59.

Upshur, J. A. (1971b). Productive communication testing : a progress report. In G. Perren and J. L. Trim (Eds.), *Applications in Linguistics* (pp.435- 42). Cambridge : Cambridge University Press.

Upshur, J. A. (1979). Functional proficiency theory and a research role for language tests. In E. J. Briere and F. B. Hinofotis (Eds.), *Concepts in Language Testing : Some Recent Studies* (pp.75-100). Washington, DC : TESOL.

Valette, R. M. (1977). *Modern Language Testing.* 2nd Ed. New York : Harcourt Brace Jovanovich.

van Ek, J. A. and Alexander, L. G. (1975). *Threshold Level English.* Oxford : Pergamon.